This document is one chapter from the EPA "Handbook for Developing Watershed Plans to Restore and Protect Our Waters," published in October 2005. The reference number is EPA 841-B-05-005. You can find the entire document http://www.epa.gov/owow/nps/pubs.html.

Handbook for Developing Watershed Plans to Restore and Protect Our Waters

Chapter 7. Analyze Data to Characterize the Watershed and Pollutant Sources

October 2005

7. Analyze Data to Characterize the Watershed and Pollutant Sources

Handbook Road Map

Chapter Highlights

- Identifying locations of impairments and problems
- Determining timing of impairments and problems
- Identifying potential sources
- Determining areas for quantifying source loads

Read this chapter if...

- You want to satisfy element *a* of the section 319 guidelines—identification of causes and sources that need to be controlled
- You want to characterize the general environmental conditions in your watershed
- You're not sure what types of data analyses you should use
- You want to conduct a visual assessment as part of your data analysis
- You want to link your analysis results with the causes and sources of pollutants in the watershed
- If you want to identify critical areas in the watershed that will need management measures to achieve watershed goals

7.1 Analyze Data to Identify Pollutant Sources

Chapter 5 discussed the first step of the watershed characterization process—identifying and gathering available data and information to assess the watershed and create a data inventory. Chapter 6 discussed the next step—conducting a preliminary data review, identifying any data gaps, and then collecting additional data if needed. All of this information will now be used in the next step—data analysis to characterize the watershed. This analysis supports the identification of watershed pollutant sources and causes of impairment, which is essential to defining watershed management needs. This chapter highlights the types of data analyses commonly used to characterize water quality and waterbody conditions and to identify watershed sources contributing to impairments and problems.

This phase of the watershed planning process should result in the first of the nine elements that EPA requires in a section 319-funded watershed plan. Element *a* is *"Identification of causes and sources or groups of similar sources that need to be controlled to achieve load reductions, and any other goals identified in the watershed plan*."

Remember that data gathering and analysis is an ongoing, iterative process. Data examined in this phase will continue to be used in subsequent activities such as identifying and evaluating management measures and tracking implementation efforts.

7.1.1 Focus Your Analysis Efforts

Although many techniques are described in this chapter, you will likely choose only a selected combination of the techniques in your watershed. The process of conducting data analyses to characterize your watershed and its pollutant sources begins with broad assessments such as evaluating the averages, minimums, and maximums of measured parameters at all watershed stations. The analyses are then systematically narrowed, with each step building on the results of the previous analysis. Through careful analysis you'll obtain a better understanding of the major pollutant sources, the behavior of the sources, and their impacts on the waterbodies. An understanding of the watershed conditions and sources is also the basis for determining the appropriate method for quantifying the pollutant loads.

In addition, the kinds of data analyses you perform will be determined by the amount of available data. For example, if you have data for several stations in a watershed, you'll be able to evaluate geographic variations in water quality throughout the watershed—an analysis you could not do with data for only one station.

Table 7-1 provides examples of data analysis activities and the tools used in various steps of the watershed planning process. It gives you an idea of how the parameter or analytical techniques might vary depending on where you are in the process and your reasons for analysis.

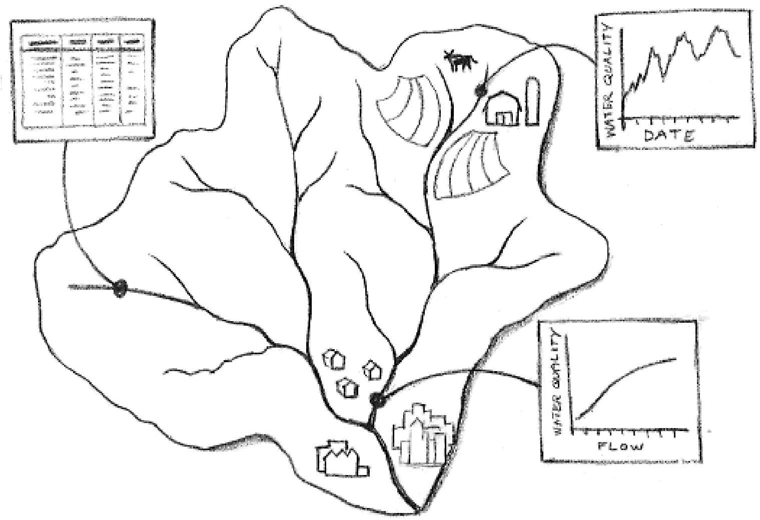
Table 7-1. Examples of the Types of Data-related Activities Conducted Throughout the Watershed Planning Process

Watershed Planning Step	Type of Data	Goal of Data Analysis	Example Activity
Characterize Watershed	• Previously conducted studies (e.g., TMDLs, 305(b) report, USGS water quality reports, university studies)	Generally characterize the watershed and identify the most important problems for further analysis.	• Review available reports and assessments.
	• Watershed data (e.g., land use, soils) • Chemical instream data • Biological instream data • Physical data	Perform targeted analysis of available data to characterize the waterbody and watershed. Examples: • Identify sources • Characterize the impairment • Evaluate spatial trends • Evaluate temporal trends • Identify data gaps	• Compare data to water quality standards to identify timing and magnitude of impairment. • Review monthly statistics to identify seasonal variations. • Use GIS at watershed stations to identify spatial variations in water quality and potential sources of pollutants.
Set Goals and Identify Solutions	Watershed data (e.g., land use, soils, population) • Chemical instream data • Biological instream data • Physical data • Meteorological data	Appropriately represent watershed and waterbody in the model for the most accurate simulation of watershed loads.	• Use data to establish a non-modeling analysis (e.g., use observed data to establish a spreadsheet mass balance calculation). • Use data for model setup (e.g., identify appropriate model parameter values, establish watershed characteristics such as land use and soils). • Compare observed data to model output for calibration and validation.
Implement and Evaluate	Instream monitoring data for the parameters of concern (e.g., nutrients)	Evaluate the effectiveness of management measures and track the progress of water quality improvement.	• Compare data collected upstream and downstream of management practices. • Compare data collected before and after implementation of management practices to track water quality improvement.

Note: TMDL = Total Maximum Daily Load; USGS = U.S. Geological Survey; GIS = geographic information system.

7.1.2 Use a Combination of Analysis Types

Because data analysis techniques are used to support a variety of goals and involve multiple types of data, a combination of techniques is usually used. Less-detailed analyses, such as evaluating summary statistics, might be conducted for certain pollutants, whereas more detailed analyses might be conducted for others, depending



on the goals of the plan and the pollutants of concern. Data analysis is typically an iterative process that is adapted as results are interpreted and additional information is gathered.

7.1.3 Consider Geographic Variations

The kinds of analyses and the level of detail used in your data analysis will vary within the watershed depending on the pollutants of concern. For example, if bacteria loading from livestock operations is a primary concern in the watershed, detailed land use analysis might be necessary to identify pasturelands and evaluate proximity to streams and water access for livestock, as well as to identify and characterize areas of cropland that receive manure applications. In addition, detailed water quality analyses might be needed for the areas that contain livestock to evaluate the timing and magnitude of impacts as related to livestock grazing schedules and access to waterbodies. For other areas of the watershed, general water quality characterization will be sufficient, and low-level evaluations of stream characteristics, watershed soils, and other types of data will be acceptable given the focus of the data analysis.

7.1.4 Incorporate Stakeholders' Concerns and Observations

Stakeholder concerns and goals will also help to determine what kinds of analyses are needed. If the stakeholders and the earlier characterization identified bacteria- and metals-associated impacts from developed areas as a primary concern, the data analysis will focus on characterizing those parameters and the locations, types, or timing of pollutant loading from urban and residential sources in the watershed. If a specific source is expected to be contributing to water quality problems, more detailed analyses might be conducted on data collected upstream and downstream of that source, or smaller time scales (e.g., daily concentrations) might be evaluated. Data analysis in the remainder of the watershed would be more coarse, identifying simple summary statistics (e.g., monthly minimum, maximum, average) sufficient for general characterization of identified subwatersheds. Table 7-2 illustrates this concept with examples of different levels of effort for the various types of data used in watershed characterization. Other factors to consider regarding level of detail, include: relative costs of remediation, risks to human health and aquatic life, and level of disagreement among stakeholders–all of which would likely increase the level of detail needed.

Table 7-2. Examples of the Level of Detail and Effort for Typical Types of Data

Type of Data	➡ Increasing level of complexity ➡		
	Low	**Moderate**	**High**
Instream (e.g., water quality, flow)	Summary statistics (e.g., minimum, average, maximum) for watershed stations	Spatial analysis of water quality using instream water quality data and GIS coverages	Spatial, temporal analysis of multiple instream parameters and GIS data (often combined with modeling and supplemental monitoring)
Land use	General distribution of land use types throughout the watershed, using broad categories (e.g., agriculture, urban)	Specific identification of land use areas by subwatershed, including more detailed categories (e.g., cropland, pasture, residential, commercial)	Statistical analysis of land use areas in relation to water quality conditions (e.g., regression analysis between amount of impervious area and average flow or water quality concentration)
Soils	General distribution of soil types based on available information	GIS analysis of the locations and types of soil series	Detailed analysis of soil distribution, including identification of proximity to streams, erosion potential, and other soil characteristics affecting soil erosion and transport

Once the focus of the data analysis has been identified, the relevant data are compiled and analyses are conducted. The following sections discuss the typical types of data analyses used to support watershed characterization and the primary data analysis techniques available to evaluate the watershed and identify causes and sources.

7.2 Analyze Instream and Watershed Data

Data analysis helps to evaluate spatial, temporal, and other identifiable trends and relationships in water quality. Analysis of instream data is needed to identify the location, timing, or behavior of potential watershed sources and their effect on watershed functions such as hydrology, water quality, and aquatic habitat. You developed a preliminary assessment of the watershed during the first and second phases of watershed characterization. Now, with a more comprehensive dataset, you can perform a more detailed and definitive analysis. One way to organize and focus the data analysis is to consider the specific watershed characteristics and the questions that need to be answered before an appropriate management strategy can be developed. ✎Use worksheet 7-1 to help determine the types of analyses you might need to conduct. ☛A blank copy is provided in appendix B.

Typical analyses used to address these questions include statistical analysis, spatial analysis, temporal analysis, trends and relationships, and flow and load duration curves. It’s important to note that most of the analyses discussed in this section focus on water quality monitoring data because many watershed goals can be directly or indirectly linked to instream water quality conditions. In addition, water quality is an indicator of the general watershed conditions and pollutant source types, locations, and behavior. However, you should also broaden the evaluation of watershed conditions by incorporating additional data types (e.g., land use, weather, and stream

✎ WORKSHEET 7-1
What Data Analyses Do We Need to Conduct?

Questions to Help Determine What Kinds of Data Analyses Are Needed

Question	Section to refer to for assistance
1. Are water quality standards being met? If so, are they maintaining existing levels?	7.2.1 (Confirm Impairments) 7.2.2 (Summary Statistics)
2. Is water quality threatened?	7.2.1 (Confirm Impairments) 7.2.2 (Summary Statistics)
3. Is water quality impaired?	7.2.1 (Confirm Impairments) 7.2.2 (Summary Statistics)
4. Are there known or expected sources causing impairment?	7.2.7 (Visual Assessment)
5. Where do impairments occur?	7.2.3 (Spatial Analysis)
6. When do the impairments occur? Are they affected by seasonal variations?	7.2.4 (Temporal Analysis)
7. Under what conditions (e.g., flow, weather) are the impairments observed?	7.2.4 (Temporal Analysis) and 7.2.5 (Other Trends and Patterns)
8. Do multiple impairments (e.g., nutrients and bacteria) coexist?	7.2.5 (Other Trends and Patterns)
9. Are there other impairments that are not measured by water quality standards?	7.2.6 (Stressor Identification)

Questions to answer based on results of the data analysis:

1. What beneficial uses for the waterbodies are being impaired? What pollutants are impairing them?
2. What are the potential sources, nonpoint and point, that contribute to the impairment?
3. When do sources contribute pollutant loads?
4. How do pollutants enter the waterbody (e.g., runoff, point sources, contaminated ground water, land uses, ineffective point source treatment, pipe failures)?
5. What characteristics of the waterbody, the watershed, or both could be affecting the impairment (e.g., current or future growth, increased industrial areas, future NPDES permits, seasonal use of septic systems)?
6. Revisit the conceptual model showing the watershed processes and sources, and revise it if necessary.

morphology) discussed in chapter 5, as necessary or appropriate for your watershed. A summary of the various types of analyses used in a watershed characterization is provided below.

7.2.1 Confirm Impairments and Identify Problems

The first step in characterizing a watershed involves understanding the water quality impairments and designated use impacts occurring in your watershed. The following reports and databases are available to support this activity:

- **305(b) report (as part of the Integrated Report)**—summarizes designated use support status for waters in the state
- **303(d) lists (as part of the Integrated Report)**—identify waters not meeting water quality standards
- **EPA's Assessment Database (ADB)**—includes data used in 305(b) and 303(d) assessments
- **TMDL Tracking System (stand-alone or through WATERS)**—includes locations of 303(d)-listed waterbodies and provides downloadable geographic information system (GIS) coverages

EPA's Assessment Database

EPA's new Assessment Database (ADB) application provides a framework for managing water quality assessment data. The ADB is designed to serve the needs of states, tribes, and other water quality reporting agencies for a range of water quality programs (e.g., CWA sections 305(b), 303(d), and 314). The ADB stores assessment results related to water quality standards designated use attainment, the pollution associated with use impairments, and documentation of probable pollution sources. The ADB can be used to generate several pre-formatted reports, as well as conventional data tables and lists. ☛For more information on using the ADB, go to www.epa.gov/waters/adb. The most recent EPA Integrated Report guidance includes an increased emphasis on using the ADB to meet reporting requirements.

Although these references provide the necessary information to *identify* the types of water quality problems occurring in your watershed, it's likely that you'll have to analyze the available monitoring data yourself to fully *characterize and understand* the problems. This analysis typically involves comparing available monitoring data to water quality standards, but in a way that goes beyond the assessment already completed by the state for section 303(d) and 305(b) assessments. When identifying "impaired" waterbodies for the 303(d) list, states usually compare available monitoring data to applicable water quality criteria and, based on their listing guidelines and criteria (e.g., percent of samples above the criteria), determine which waters don't meet the criteria. In evaluating impairments in your watershed, you don't want to simply duplicate the state's efforts. Instead, use the 305(b) and 303(d) information to target your analyses—to identify which waterbodies are impaired or threatened—and begin your analysis there. (You should also include in your analysis those waterbodies identified by stakeholders as degraded but not included in the state assessments.)

It's a good idea to do a general analysis (e.g., summary statistics) of all the waterbodies and associated data in your watershed, but you can focus the more in-depth evaluation of impairment on those waterbodies known to have problems. To better understand the watershed impairments, you can analyze the water quality and instream data in a variety of ways. The first likely analysis is simply the magnitude of the impairment—how bad is the problem? Identifying the percentage of samples that violate standards provides insight into the level of impairment in the watershed, or at a particular location. Using a graphical display of water quality data compared to applicable criteria is also an easy way to generally illustrate the frequency and magnitude of standards violations, as shown in figure 7-1. A temporal analysis of water quality versus standards can be used to identify the times of year, season, month, and even day when the impairment is occurring or is the worst. Temporal and other analyses are discussed further in this section. These analyses are used to understand the general watershed conditions and to support identification of pollutant sources, but also provide information specific to the distribution, timing and magnitude of water quality impairment.

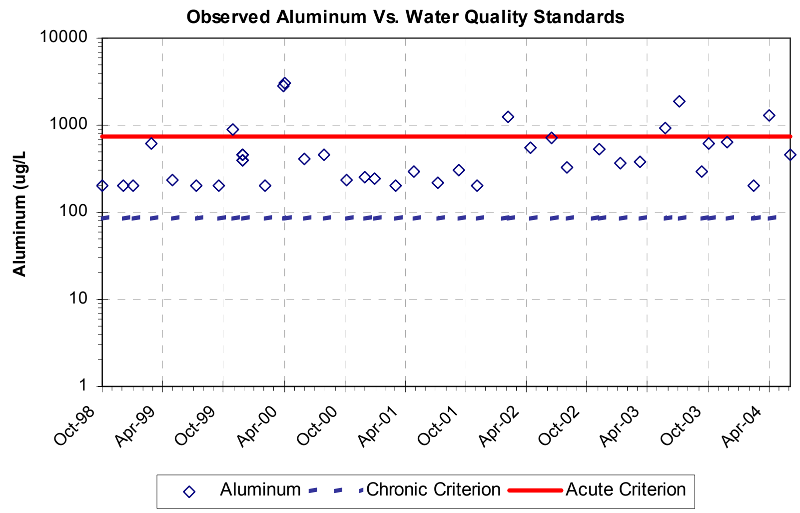


Figure 7-1. Example graph of observed aluminum concentrations compared to water quality criteria.

7.2.2 Summary Statistics

Statistical analyses are essential tools for describing environmental data and evaluating relationships among different types of data. You might not need to conduct in-depth statistical testing to characterize your watershed, but it's often useful to develop summary statistics to summarize your available datasets, to help in preliminary analysis, and to communicate your results to stakeholders and the public. Summary statistics include such characteristics as range (e.g., minimum, maximum), central tendency (e.g., mean, median), and variability (standard deviation, coefficient of variation). Figure 7-2 defines many of the commonly used statistical terms. Summary statistics should be computed for all stations and relevant data (e.g., pollutants of concern) as one of the first steps in your data analysis. Microsoft Excel and other spreadsheet programs make developing summary statistics simple. The program can automatically calculate any of the statistical functions based on the dataset. In addition, you can create Pivot tables in Excel that calculate several statistical functions for any combination of the data at once (e.g., by pollutant by station). It is useful to also calculate the number or percentage of

More on Statistics

This section discusses the typical types of data analyses to support watershed characterization and identification of pollutant sources. Each analysis can be conducted with varying degrees of detail and complexity. In addition, it might be useful to perform more detailed statistical tests. For example, a Mann-Kendall test can be applied to long-term datasets to indicate whether there is a statistically significant increasing or decreasing trend in the water quality data. Available references with information on statistical analysis of environmental data include

Helsel, D.R., and R.M. Hirsch. 2002. Statistical Methods in Water Resources. Chapter A3 in Book 4, *Hydrologic Analysis and Interpretation*, of Techniques of Water-Resources Investigations of the United States Geological Survey. http://water.usgs.gov/pubs/twri/twri4a3

NRCS (Natural Resources Conservation Service). 1997. *National Handbook of Water Quality Monitoring*. 450-vi-NHWQM. National Water and Climate Center, Portland, Oregon.

Measures of Range: Identify the span of the data from low to high.

Minimum: The lowest data value recorded during the period of record.

Maximum: The highest data value recorded during the period of record.

Measures of Central Tendency: Identify the general center of a dataset.

Mean: The sum of all data values divided by the sample size (number of samples). Strongly influenced by outlier samples (i.e., samples of extreme highs or lows); one outlier sample can shift the mean significantly higher or lower.

Median ($P_{0.50}$): The 50^{th} percentile data point; the central value of the dataset when ranked in order of magnitude. The median is more resistant to outliers than the mean and is only minimally affected by individual observations.

Measures of Spread: Measure the variability of the dataset.

Sample variance (s^2) and its square root standard deviation (s): The most common measures of the spread (dispersion) of a set of data. These statistics are computed using the squares of the difference between each data value and the mean, and therefore outliers influence their magnitudes dramatically. In datasets with major outliers, the variance and standard deviation might suggest much greater spread than exists for most of the data.

Interquartile range (IQR): The difference between the 25^{th} and 75^{th} percentile of the data. Because the IQR measures the range of the central 50 percent of the data and is not influenced by the 25 percent on either end, it is less sensitive to extremes or outliers than the sample variance and standard deviation.

Measures of Skewness: Measures whether a dataset is asymmetric around the mean or median and suggests how far the distribution of the data differs from a normal distribution.

Coefficient of skewness (g): Most commonly used measure of skewness. Influenced by the presence of outliers because it is calculated using the mean and standard deviation.

Quartile skew coefficient (qs): Measures the difference in distances of the upper and lower quartiles (upper and lower 25 percent of data) from the median. More resistant to outliers because, like the IQR, uses the central 50 percent of the data.

Figure 7-2. Commonly used summary statistics.

samples violating water quality criteria to include in your summary statistics for each station.

7.2.3 Spatial Analysis

If evaluation of the summary statistics for the water quality stations in your watershed indicates noticeable differences in water quality throughout the watershed, you should do a more focused analysis of spatial variation in water quality and other waterbody monitoring data. Spatial analysis of available waterbody data can be useful to

- Determine the general distribution of water quality or habitat conditions
- Identify the locations of areas of concern or potential major sources
- Determine the impact of a specific source
- Identify the effect of a management practice or control effort

The spatial distribution of water quality conditions in the watershed might indicate the location of "hot spots" and sources potentially affecting impairment. Spatial analysis of data is also useful in evaluating the potential impacts of specific sources, when sufficient data are available. Evaluating the difference in paired observations from stations upstream and downstream of a potential source can indicate the impact of the source on instream conditions. Similar data analysis can be conducted on data available upstream and downstream of a management practice to evaluate the effectiveness of the management practice in reducing pollutant loads to the waterbody.

Simply reviewing a table of summary statistics for each station in the watershed can identify areas of varying water quality. When dealing with a large watershed with multiple stations, however, a GIS can be used to effectively present and evaluate spatial variations in water quality conditions, as shown in the example map in figure 7-3. Presenting water quality summaries by station throughout a watershed in GIS also allows for identification of corresponding watershed conditions or sources that might be causing the spatial variations, such as land use distribution and location of point sources. This information is important for identifying the potential sources that might be causing the watershed problems and impairments.

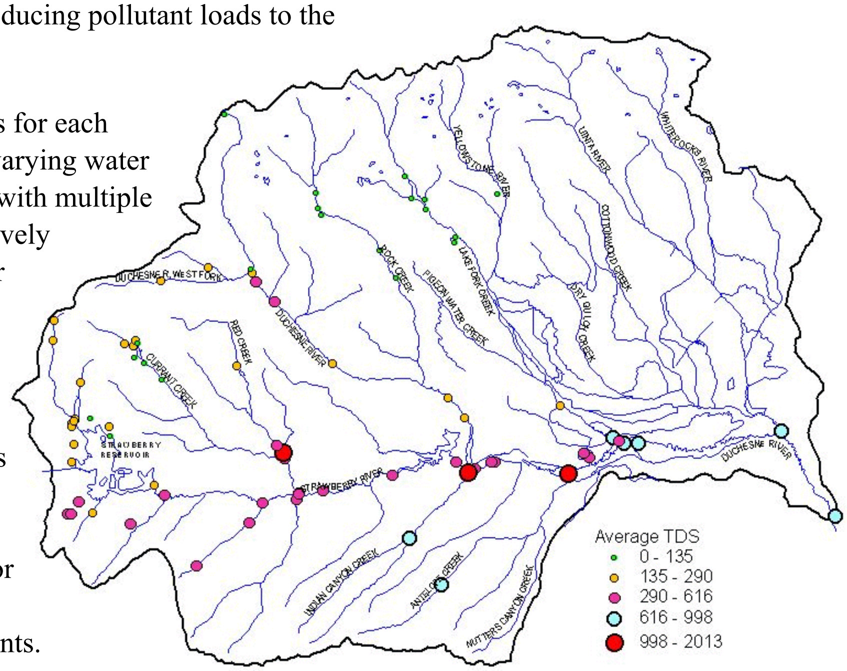


Figure 7-3. Example map of average total dissolved solids concentration throughout a watershed.

Even if sufficient monitoring data are not available to adequately evaluate spatial variation in water quality, you should still evaluate other available watershed data to understand the spatial distribution of characteristics that are likely influencing waterbody conditions, such as land use, soils, and location of permitted sources. GIS is a very useful tool for displaying and evaluating these kinds of data.

7.2.4 Temporal Analysis

Another important analysis is the evaluation of temporal trends in water quality conditions. Evaluation of temporal patterns can assist in identifying potential sources in the watershed, seasonal variations, and declining or improving water quality trends. Temporal analyses can include long-term trend analysis to identify generally increasing or decreasing trends in data and more focused analysis of monthly, seasonal, and even daily and hourly variations.

Degraded water quality during certain months or seasons can indicate the occurrence of a source that is active only during those times. For example, elevated concentrations of nutrients or bacteria during the summer months (figure 7-4) might indicate increased source activity, such as livestock grazing, during those months. It might also indicate a need for further analysis of other watershed conditions (e.g., weather, flow) that can exacerbate the impairment during the summer months. For example, warmer temperatures during the summer might increase the productivity of algae, leading to greater decreases in dissolved oxygen.

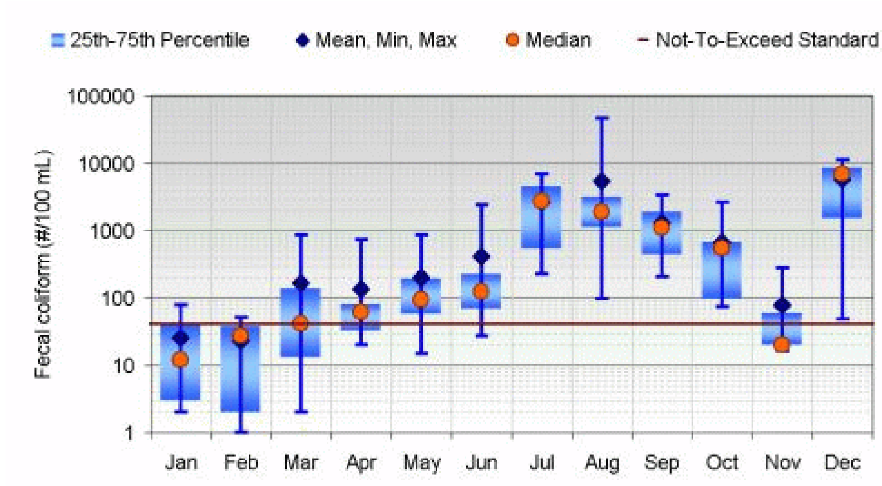


Figure 7-4. Example graph of monthly statistics for fecal coliform bacteria.

7.2.5 Other Trends or Patterns

It is often beneficial to evaluate relationships and trends in the available data other than spatial and temporal trends. Important examples include

- Evaluating the relationship between flow and instream water quality (☛see chapter 5 for data sources)
- Documenting the relationship between related pollutants
- Evaluating the relationship of instream conditions to other watershed factors (e.g., land use, source activity)

Flow Versus Water Quality

An identifiable relationship between flow and instream water quality concentrations can indicate what types of pollutant sources dominate the instream impairment and can help to identify critical conditions surrounding the impairment. For example, runoff-driven nonpoint sources typically dominate instream water quality conditions during periods of high flow resulting from rainfall/runoff events, whereas point sources that provide relatively constant discharges to receiving waters usually dominate water quality during low flow, when there is less water to dilute effluent inputs.

There are several options for evaluating the relationship between flow and a water quality parameter, including visually evaluating time series data, developing a regression plot, evaluating monthly averages, and developing a flow duration curve.

Using Duration Curves to Connect the Pieces

America's Clean Water Foundation published an article discussing duration curves and their use in developing TMDLs (Cleland 2002). The duration curves act as an indicator of relevant watershed processes affecting impairment, important contributing areas, and key delivery mechanisms. ☛To read the full article and get more information on the use of duration curves to diagnose seasonal impacts and potential sources, go to www.tmdls.net/tipstools/docs/BottomUp.pdf.

A flow duration curve can be a useful diagnostic tool for evaluating critical conditions for watershed problems and the types of sources that could be influencing waterbody conditions. Flow duration curves graph flows based on their occurrence over the period of record. Flows are ordered according to magnitude, and then a percent frequency is assigned to each, representing the percentage of flows that are less than that flow. For example, a flow percentile of zero corresponds to the lowest flow, which exceeds none of the flows in that record. The percentage of 100 corresponds to the highest flow, which exceeds all the flows in that record. The flow duration is often plotted with corresponding pollutant concentrations to evaluate the relationship between water quality and flow. To do this, you should isolate matching flow and water quality and plot the flow and concentration data as a function of flow percentile.

A variation of the flow duration curve is the load duration curve, which plots observed pollutant loads as a function of flow percentile. Matching water quality and flow (measured on the same day) are used to calculate observed loads, by multiplying flow by pollutant concentration and an appropriate conversion factor. The loads are then plotted along with the flow in order of flow percentile. The load duration curve provides information on when loading occurs.

As shown in the example load duration curve (figure 7-5), the total dissolved solids (TDS) concentrations tend to follow a pattern similar to the flow, with lower concentrations occurring during lower flows and elevated concentrations during higher flows. This indicates that surface runoff (nonpoint sources or stormwater discharges) is likely the source of elevated total dissolved solids rather than point source discharges. The flow duration method does not allow you to identify specific sources (e.g., residential versus agricultural), but it provides useful information on the conditions under which problems occur and the general types of sources affecting the waterbody.

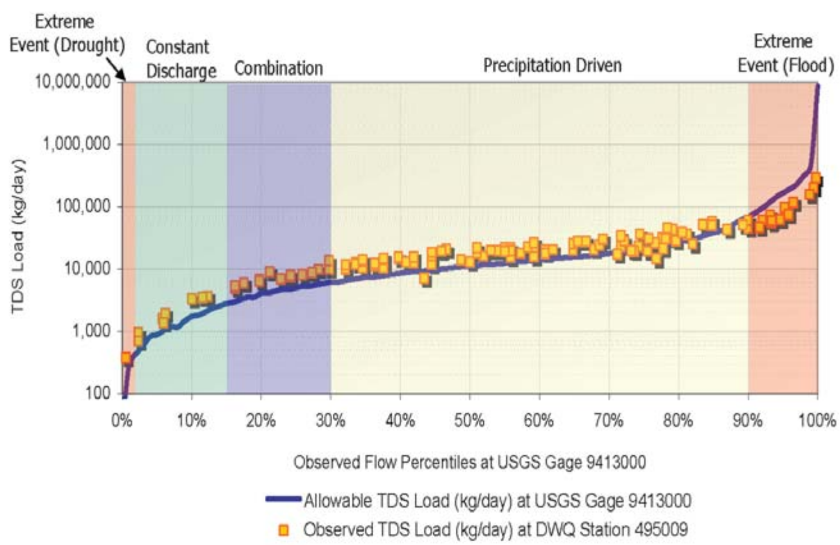


Figure 7-5. Example load duration curve.

Relationships Between Pollutants

It's also important to evaluate the correlation of instream concentrations (and loading) of pollutants of concern to other parameters that represent the same impairment or are likely being contributed by similar sources or acting as a source of the pollutant of concern. For example, metals often attach to sediments, resulting in increased metals loading during times of high sediment erosion and runoff.

Establishing a correlation between instream sediment and metal concentrations can indicate that metals loading in the watershed is sediment-related. Understanding these relationships will be important when establishing load reductions and selecting appropriate management activities.

Using the Correlation of Phosphorus, pH, and Chlorophyll a to Understand Instream Conditions and Focus Management Efforts

The Vandalia Lake, Illinois, TMDL establishes load reduction goals for total phosphorus to address impairments from both phosphorus and pH. Fluctuations in pH can be correlated to photosynthesis from algae. Chlorophyll a indicates the presence of excessive algal or aquatic plant growth, which is a typical response to excess phosphorus loading. Reducing total phosphorus is expected to reduce algal growth, thus resulting in attainment of the pH standard. Available monitoring data for the lake were used to evaluate the relationship between pH, chlorophyll a, and total phosphorus. The general relationships suggested that controlling total phosphorus will decrease chlorophyll *a* concentrations, which will in turn reduce pH into the range required for compliance with water quality standards.
☛www.epa.state.il.us/water/tmdl/report/vandalia/vandalia.pdf

Waterbody Conditions Versus Watershed Characteristics

Evaluating relationships between instream conditions and watershed features or conditions will also facilitate the identification of sources and the establishment of successful management goals and focused implementation efforts. For example, performing statistical analyses on instream data and watershed features, such as weather patterns, land use (e.g., percent impervious, area of urban), or soils (e.g., erodibility), can establish a quantitative link between watershed conditions and the resulting instream conditions. It might also be appropriate to divide data into separate datasets representing certain time periods or conditions for evaluation (e.g., storm event vs. base flow, irrigation season, grazing season).

7.2.6 Stressor Identification

When waterbodies experience biological impairment due to unknown causes, stressor identification is used to identify the most likely causes of biological impairment (figure 7-6). This formal method of causal evaluation can be used in a number of ways:

- To increase confidence that costly remedial or restoration efforts are targeted at factors that can truly improve biological condition
- To identify causal relationships that are otherwise not immediately apparent
- To prevent biases or lapses of logic that might not be apparent until a formal method is applied

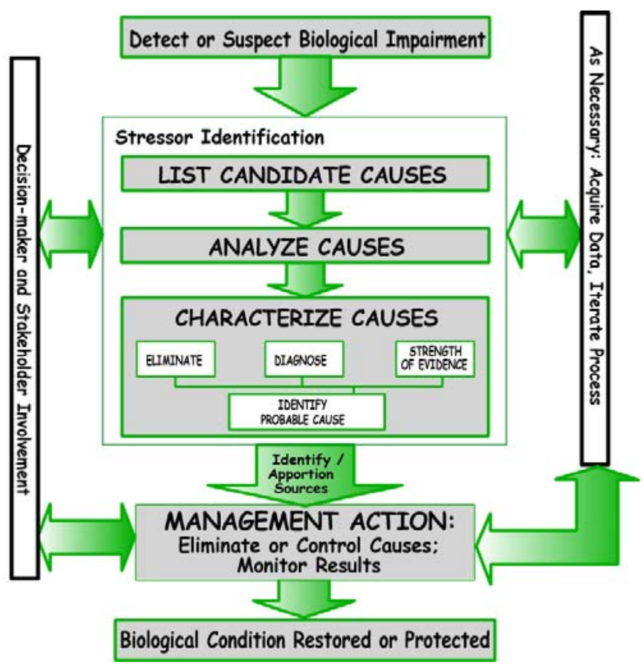


Figure 7-6. Stressor identification process.

☛For a detailed description of the stressor identification process, see EPA's *Stressor Identification Guidance Document* (USEPA 2000b; www.epa.gov/waterscience/biocriteria/stressors/stressorid.html). In addition, two stressor identification modules originally developed as part of EPA's 2003 National

Biocriteria Workshop are available online. ☛The SI 101 course contains several presentations on the principles of the stressor identification process: www.epa.gov/waterscience/biocriteria/modules/#si101.

EPA recently released the Causal Analysis/Diagnosis Decision Information System (CADDIS) to support determination of causes of biological impairment. CADDIS is an online tool that helps investigators in the regions, states, and tribes to find, access, organize, use, and share information to produce causal evaluations of aquatic systems. It is based on the EPA's stressor identification process. Current features of CADDIS include

- Step-by-step guide to conducting a causal analysis
- Downloadable worksheets and examples
- Library of conceptual models
- Links to helpful information

☛Go to the CADDIS Web site at http://cfpub.epa.gov/caddis/home.cfm to access CADDIS and obtain more information.

7.2.7 Visual Assessments and Local Knowledge

It's important to remember that monitoring and GIS data can provide only a representation of your watershed. Depending on the frequency of monitoring, the data might not reflect chronic conditions but rather provide a snapshot of conditions unique to the time of sampling, especially when dealing with parameters that are highly variable and sensitive to localized impacts (e.g., bacteria counts). To make the most of your data analysis, it is important to analyze the data with an understanding of the "real world." Use the data analysis to support what you already know about the watershed from the people that live and work there. ☛As discussed in sections 4.3.2 and 6.5.1, visual assessments (e.g., streamwalks, windshield surveys) are useful for identifying and connecting potential sources of impairment and watershed conditions and should be used to guide and support data analysis for identifying watershed sources. In watersheds with limited monitoring data, visual assessments are especially important, providing the basis for source identification.

Not only are visual assessments useful for identifying potential pollutant sources and areas to focus your data analysis, but they can also answer questions raised by your data analysis. For example, if your data analysis shows a dramatic decrease in water quality in a portion of your watershed, but the land use and other watershed coverages don't indicate any major sources in that area, it's a good idea to walk the stream or drive through the area to identify any possible reasons for the change. For example, your data might indicate sharp increases in sediment measures (e.g., turbidity, total suspended solids) between two monitoring stations. Reviewing the land use maps

Examples of Sources You Might Miss without a Watershed Tour

- Streambank erosion
- Straight pipes
- Livestock (near or with access to streams)
- Wildlife (e.g., waterfowl populations on lakes and open streams)

do not suggest any activities that would account for such a dramatic increase. When you drive through the watershed, you might find a source that you would never know about without surveying the area, such as a severely eroding streambank or livestock or wildlife watering in the stream and causing resuspension of streambed sediments.

In addition to visual inspection of the watershed, local knowledge and anecdotal information from stakeholders are often very important to successfully analyzing and interpreting your watershed data. They, too, can provide useful insight to support or guide data analysis, especially if it is historical information that would not be identified through a present-day visual assessment. A data analysis conducted for Lake Creek, Idaho, provides an example of stakeholder anecdotal information's being crucial to identifying a watershed source. The data analysis indicated an unexplained increase in turbidity and sediment between two stations in the stream (figure 7-7). Discussing the data analyses with stakeholders allowed TMDL developers to understand that the increase was the result of localized logging that had occurred near the stream several years earlier. Knowing that the logging had occurred explained why the turbidity levels had dramatically and quickly increased at the downstream station and were now still recovering. Without this knowledge, the TMDL might have inappropriately targeted areas that were not affecting the stream.

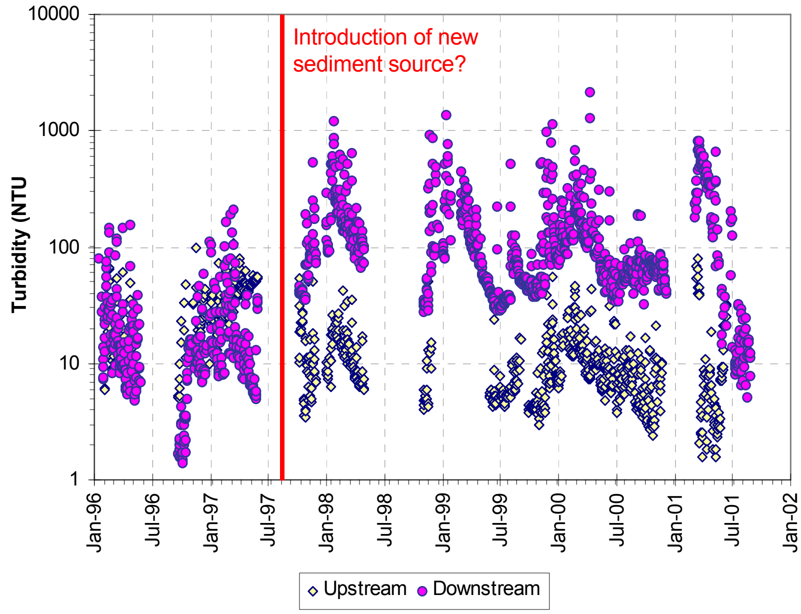


Figure 7-7. Long-term turbidity levels at two stations in Lake Creek, Idaho.

7.3 Evaluate Data Analysis Results to Identify Causes and Sources

Together with the input from stakeholders and your local knowledge of the watershed, analyzing your data should lead you to an understanding of where and when problems occur in your watershed and what could be causing the problems. Ideally the data analysis phase will progress in such a manner that each analysis leads to a greater understanding of the problems, causes, and sources—each analysis identifies another piece of the puzzle. Suppose, for example, that you started your analysis with a calculation of summary statistics for bacteria at all the stations in your watershed. In doing so, you noticed that stations in the upstream portion of the watershed had higher averages, maximums, and minimums than the rest of the watershed. Focusing on those stations, you began to evaluate temporal variations, noting that bacteria levels were consistently higher during the spring and summer. From there you began to look at other factors that might change seasonally, including weather, flow, and surrounding land activities. You discovered that although rainfall and flow are higher during the spring, possibly delivering higher bacteria loads, they are lower during the summer. Also, rainfall and flow are higher throughout the watershed, not in only this "problem area." So, what else might be causing the higher levels during those two seasons? By evaluating land use data for the surrounding area, you realize there are some concentrated pockets of agricultural land in the area. After talking to stakeholders and driving the watershed, you identify several acres of pastureland used for horse and cattle grazing during the spring and summer. Much of the pastureland is in close proximity to the streams with elevated observed bacteria, and some of the pastures even have direct access to the streams. Such a combination of focused data analyses, visual assessments, and local knowledge is critical to identifying and understanding watershed sources.

In addition, the data analysis will identify on which sources you'll need to focus during the loading analysis discussed in chapter 8. Some sources will be expected to have a greater impact on watershed problems than others and might require more detailed analysis. For example, if runoff from developed areas is expected to be the primary cause of elevated metals in watershed streams, it might not be necessary to evaluate subcategories of agricultural or other undeveloped lands in the loading analysis. You can likely group those land uses or sources together and focus on the developed areas, possibly even breaking them into more detailed categories (e.g., suburban, commercial).

7.3.1 Grouping Sources for Further Assessment

Once you understand the potential causes and sources of the watershed problems, you should decide at what level you want to characterize those sources. The next step of the process is to quantify the watershed sources—to estimate the pollutant loads contributed by the sources (chapter 8). Therefore, you should identify the sources you want to quantify. The level of detail in estimating the source loads can vary widely and will depend largely on the results of your data analysis. The analysis should give you an understanding of the sources that are affecting watershed and waterbody conditions, providing a guide for which sources need to be controlled. Therefore, it's important to identify sources at a level that will result in effective control and improvement. For example, if you have identified specific pastures in one portion of the watershed as dominating the bacteria levels in your watershed during the summer, it would not be appropriate to quantify agricultural or even pastureland sources as an annual gross load for the *entire* watershed.

Example Categories for Grouping Pollutant Sources

- Source type (e.g., nonpoint, point)
- Location (e.g., subwatershed)
- Land use type
- Source behavior (e.g., direct discharge, runoff, seasonal activities)

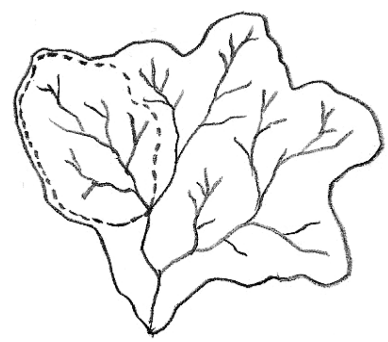
To facilitate estimation of source loads, and later source control, sources should be grouped into logical categories that help to prioritize and address certain pollutants, sources, or locations for more efficient and effective management. Consider the following factors and methods when grouping sources for assessment. You can combine many of the methods to create various groupings and layers of sources, relevant to the needs and priorities of the watershed-based plan.

Nonpoint Source Versus Point Source

Although watershed plans typically focus on nonpoint sources, they should consider and integrate point sources for effective watershed protection. You should separate nonpoint sources from point sources for assessment for both technical and programmatic reasons. Nonpoint and point sources typically behave differently and affect the receiving waters under different conditions. For example, nonpoint sources usually contribute pollutant loads that are washed off and transported during precipitation events, affecting waterbody conditions during times of higher surface runoff and, therefore, higher flow. Point sources usually discharge constant loads to receiving waters, affecting waterbody conditions during times of low flow when there is less water to dilute incoming effluents. Not only do point and nonpoint sources behave and affect waterbodies differently, but their management and control mechanisms are also different. Grouping them separately when considering future implementation of control measures is logical.

Spatial Distribution and Location

Grouping sources by location facilitates their assessment by dividing the area of concern into smaller, more focused areas, and it often supports future implementation. Spatially grouping sources helps to identify priority regions or locations that should be targeted for control. The method of grouping sources typically involves creating subwatersheds within the larger watershed of concern and also prioritizing sources within the subwatershed by some other methodology (e.g., proximity to a stream, land use).



Land Use Distribution

Sources are often specific to certain land uses, making it logical to group them by land use. For example, sources of nutrients such as livestock grazing and fertilizer application, which occur in conjunction with agricultural land use, would not likely contribute loads to other land uses such as urban or forest uses. Likewise, urban land uses typically have a set of pollutants of concern (e.g., metals, oil, sediment) different from those of rural land uses based on the active sources. Although it is difficult to isolate inputs from individual sources within a land use, assessing them as land use inputs can still support evaluation of loading and identification of future controls. Sources can be grouped and characterized by land use at a large scale, such as all agricultural lands, or at a very detailed level, such as specific crop type. In some cases, subcategories of nonpoint sources should be used to estimate the source contribution. For example, a land use like agriculture would often be further broken down into grazing or cropland, allowing a more accurate estimate of the sources coming from each subcategory and the ability to choose the most effective management practices for each subcategory.

Grouping sources according to their land use also facilitates identification of future implementation efforts because certain management practices are most effective when applied to a certain land use.

Delivery Pathway and Behavior

Depending on their behavior, nonpoint sources can contribute pollutants to receiving waters through different delivery pathways. The nature of the delivery might support separate assessment of the source. For example, grazing cattle might be treated as a separate source depending on the activity or location of the cattle. Livestock on rangeland can contribute pollutants to the land that are picked up in runoff, whereas livestock in streams deposit nutrient and bacteria loads directly to the streams. Different methods might be required to evaluate the effect of each group on waterbody conditions. Another example is failing septic systems that might be contributing pollutant loads to waterbodies. Because loads from the septic systems can be delivered through groundwater and also through surface breakouts, you might conduct separate analyses to estimate their loads.

Other Factors

Additional factors that can influence the grouping of sources include the following:

- **Social and economic factors.** Certain sources and their impact might be of higher priority to the affected public because they are more visible than other sources or because they could have negative impacts on the local economy. Public buy-in and priorities can influence the evaluation and grouping of sources, as well as subsequent source control.

- **Political jurisdictions.** Because source control can ultimately fall to different jurisdictions (e.g., counties), it might be necessary to evaluate sources based in part on jurisdictional boundaries. In some cases, the sources might even be subject to different laws and control options depending on where they're located.

7.3.2 Time Frame for Source Assessment

Another important consideration when deciding how to quantify your sources is the time frame you want to capture. Your data analysis should provide insight into the timing of watershed problems and, therefore, into the temporal scale you need to evaluate sources. For example, instream dissolved oxygen might decrease only during summer months because of increased nutrient loading, higher temperatures, and lower flows. Therefore, it will be important to characterize and quantify sources on a time scale that allows for evaluation during the summer months. It would not be appropriate to evaluate annual loading for a problem that occurs only during the summer.

7.4 Summarize Causes and Sources

On the basis of your data analysis, you should now be able to identify the key sources you will quantify in the next step of the watershed planning process. You should identify the source type, locations, and timing for load estimation (chapter 8). It might be helpful to identify the areas for evaluation on a watershed map to determine the key locations for conducting the loading analysis and which sources will be included in the analysis. You should also develop a brief report summarizing your data analyses and their results and describing the watershed sources, including their location, associated pollutants, timing, and impact on the waterbody.

In identifying your sources and grouping them for load estimation, you'll also begin to identify the critical areas needed for implementing management measures, as required as element *c* of the nine minimum elements. Element *c* is "*A description of the nonpoint source management measures that will need to be implemented to achieve load reductions and a description of the critical areas in which those measures will be needed to implement this plan.*" At this step, you have identified the recommended source groupings and priorities and you'll continue to refine the groupings as you conduct your loading analysis (chapter 8) and target your management measures (chapters 10 and 11). You'll identify the final critical areas when you select the management strategies for implementing your plan (chapter 11), but the sources and associated groupings and characteristics you have identified at this stage will provide the basis and groundwork for identifying those critical areas.